Assessing The Limits of Synthetic Controls: Estimating Causal Effects in Complex Data Structures

By Lennart Bolwin and Justus Töns*

Potential framework: We argue that applications of Synthetic Controls are faced with a self-selection problem. That is, the method is primarily applied to non-complex data structures that are straightforward to forecast (e.g. development of GDP). Using Monte Carlo studies, we show that the high interpretability of Synthetic Controls comes at the costs of poor in-sample predictions when the data structure exhibits a high degree of complexity like in stock market time series. To address this issue, we introduce the intricacy-statistic that informs the applied reseacher whether or not the data at hand exceeds the level of complexity that Synthetic Controls can handle. If the case, more flexible methodologies like the autoregressive distributed lag model can be employed to estimate the counterfactual. To do so, we introduce the R-package complex_synths that provides ready-to-use functions to compute the intricacy-statistic and, based on the magnitude of the statistics, the functionalities to estimate either the Synthetic Control or the ARDL model. To probe the performance of our methodology outside the quasi-experimental setting, we apply it to a highly complex data structure: The inclusion of a stock in an index. Specifically, we find that the inclusion of the German multi-national eCommerce company Zalando in the German stock index (DAX) caused an excess capitalization of XXX milion euro.

I. Introduction

some text

II. Research Design

some text

III. Monte Carlo Study

some text

IV. Empirical Application

some text

* Thank You Jörg!

V. Concluding Remarks

References here (manual or bibTeX). If you are using bibTeX, add your bib file name in place of BibFile in the bibliography command.

REFERENCES

Mathematical Appendix